Numerical Weather Prediction

Forecasting Research Scientific Paper No. 68

Modelling of error covariances by four-dimensional variational data assimilation

By A.C. Lorenc

June 2002

© Crown Copyright 2002

Met Office , NWP Division , Room 344, London Road , Bracknell , Berkshire ,RG12 2SZ, United Kingdom

Numerical Weather Prediction

Forecasting Research Scientific Paper No. 68

Modelling of error covariances by four-dimensional variational data assimilation

June 2002

Andrew C. Lorenc

© Crown Copyright 2002

(Submitted to the *Quarterly Journal of the Royal Meteorological Society*)

Met Office

Met Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

© Crown Copyright 2002

Permission to quote from this paper should be obtained from the Director NWP at the above address.

For additional copies, or information about other NWP publications, contact
john.armstrong@metoffice.com
Tel: 44 (0)1344 856245
Fax: 44(0)1344 854026

Modelling of error covariances by four-dimensional variational data assimilation

Andrew C. Lorenc

Summary

The extended Kalman filter is presented as a good approximation to the optimal assimilation of observations into an NWP model, as long as the evolution as errors stays close to linear. The error probability distributions are approximated by Gaussians, characterised by their mean and covariance. The full nonlinear forecast model is used to propagate the mean, and a linear model (not necessarily tangent to the full model) the covariances. Since it is impossible to determine the covariances in detail, physically based assumptions about their behaviour must be made, for instance three-dimensional balance relationships are used. The linear model can be thought of as extending the covariance relationships to the time dimension. Incremental 4D-Var is derived as a practical implementation of the extended Kalman filter, optimally using these modelled covariances for a finite time-window. It is easy to include a simplified model of forecast errors in the representation. This Kalman filter based paradigm differs from more traditional derivations of 4D-Var in attempting to estimate the mean, rather than the mode of the posterior pdf. The latter is difficult for an NWP system representing scales which exhibit chaotic behaviour over the period of interest.

The covariance modelling assumptions often result in a null space of error modes with little variance. It is argued that this is as important as the variances and correlations structures usually examined, since the implied constraints allow optimal use of observations giving gradient and tendency information.

Difficulties arise in the approach when the NWP system is capable of resolving significant structures (such as convective cells) not always determined by the observations.

Keywords: four-dimensional variational data assimilation, extended Kalman filter, numerical weather prediction, error covariance modelling

1. Introduction

When assimilating data for numerical weather prediction (NWP), imperfect observations are combined with information brought forward in time from earlier observations by the NWP model. The weights given to the latest observations should depend on the accuracy of the forecast information, which in turn depends on the position and accuracy of earlier observations. The four-dimensional variational data assimilation (4D-Var) algorithm has been developed as a practicable means of achieving this, at least for observations within a time-window. In this paper I introduce the incremental 4D-Var equations (Courtier *et al.* 1994) in a slightly unusual way, to show that the linear model and its adjoint used in incremental 4D-Var are just part of a four-dimensional covariance model. I discuss the important properties of covariances in NWP assimilation, and use analogy with dynamical relationships used for years in three-dimensional covariance models to understand the role of the linear forecast model in 4D-Var. The aim is to understand which properties of the covariances it is important to retain, when

approximating. These ideas underly the design of the 4D-Var scheme currently being developed at the Met Office (Lorenc 1997, Lorenc *et al.* 2000).

In principle, knowledge of the error probability distribution functions (pdfs) enables the optimal combination of inaccurate information. Using Bayes theorem one can derive equations for the optimal combination of new observations with a prior estimate of the atmospheric state (Lorenc 1986). The time-evolution of a pdf describing our knowledge about the atmospheric state can be calculated using the Fokker-Planck equation, also known as Kolmogorov's forward equation (Jazwinski 1970), or, ignoring model error, using the Liouville equation (Ehrendorfer 1994a&b). However in practice the numerical integration of these equations is completely impractical for a phase space with more than a few dimensions, let alone the 10^7 dimensions of a state-of-the-art NWP model. It is only practicable to propagate in time either a few characteristics, or a crude sample of the complete pdf. More importantly, as pointed out by Dee (1991), it is impossible accurately to quantify the error pdf of our forecast of the atmospheric state. To do so empirically, from data on the errors in historical forecasts, would require accurate validation data which we do not have, and more forecasts than could ever be made in the lifetime of a forecasting system.

The first approximation made by almost all practical data assimilations schemes is to model pdfs by multi-dimensional[1] Gaussian distributions. They can then be described by their mean and covariance, and equations for the optimal use of new observations become manageable (Lorenc 1986). In four-dimensional data assimilation, using a linear forecast model, the optimal sequential method is the Kalman filter (Jazwinski 1970). This is reviewed in section 2. But further approximations are needed in assimilation for NWP, because of the size and complexity of accurate forecast models. The approximations used in modelling the time behaviour of forecast error covariances in data assimilation for NWP are the theme of this paper.

The atmosphere is chaotic. Some small perturbations grow rapidly, while most perturbations are damped. The errors patterns which are similar to the growing perturbations will dominate the forecast error covariance, while others will have much smaller variance. As an analogy preparing the way for discussion of time-evolving covariances, in section 3 I discuss ways that these properties of forecast covariances can be exploited in three-dimensional variational assimilation (3D-Var).

4D-Var was introduced by Talagrand and Courtier (1991) and Lewis and Derber (1985) as a way of fitting a time-evolving trajectory from an NWP model to observations. Incremental 4D-Var (Courtier *et al.* 1994) was proposed as a practicable way for affording operational implementation. Section 4 gives a slightly unusual presentation, emphasising the equivalence with the Kalman filter by including forecast model errors. The 4D-Var equations can be expressed in the same form as 3D-Var, with time evolution within the covariance matrix. This approach shows incremental 4D-Var to be simply a way of implementing a four-dimensional covariance model. The perfect model assumption usually made when implementing 4D-Var is analogous to other assumptions such as geostrophy which can be made in modelling covariances in 3D-Var.

---

[1] Representation of the continuous atmosphere by a large but finite number of parameters is a necessary approximation for NWP. Doing this from the outset simplifies the presentation, since the number of dimensions is finite.

A totally different approach to modelling forecast covariances is the ensemble Kalman filter (Eversen 1994, Houketamer & Mitchell 1998). I discuss this from the same viewpoint, and compare with 4D-Var, in a second paper (Lorenc 2002).

2. The Kalman filter

If a pdf is Gaussian, it can be represented by its mean and its error covariance. If a variable with a Gaussian pdf is operated on by a linear operator, the output variable also has a Gaussian pdf, with error covariance determined by pre- and post-multiplying the input covariance. In these circumstances the optimal discrete sequential estimation is the Kalman filter (Jazwinski 1970). Using the notation of Ide *et al.* (1997), we denoting the model forecast state at time t_i by vector $\mathbf{x}^f(t_i)$ with error covariance $\mathbf{P}^f(t_i)$, the observations by vector $\mathbf{y}_i^o$ with error covariance $\mathbf{R}_i$, and the linear operator for calculating estimates of the observed values from a model state by $\mathbf{H}_i$. Then the posterior pdf, obtained after combining the observations and the forecast, is represented by its mean $\mathbf{x}^a(t_i)$ and its error covariance $\mathbf{P}^a(t_i)$, which are given by

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{P}^f(t_i)\mathbf{H}_i^T\left(\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i\right)^{-1}\left(\mathbf{y}_i^o - \mathbf{H}_i\mathbf{x}^f(t_i)\right) \tag{1}$$

$$\mathbf{P}^a(t_i) = \mathbf{P}^f(t_i) - \mathbf{P}^f(t_i)\mathbf{H}_i^T\left(\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}_i^T + \mathbf{R}_i\right)^{-1}\mathbf{H}_i\mathbf{P}^f(t_i) \tag{2}$$

These are forecast using a linear model $\mathbf{M}_i$, to give best estimates prior to the next batch of observations:

$$\mathbf{x}^f(t_{i+1}) = \mathbf{M}_i\mathbf{x}^a(t_i). \tag{3}$$

The discretised dynamics of the true atmosphere $\mathbf{x}^t$ is assumed to differ from (3) by stochastic perturbations:

$$\mathbf{x}^t(t_{i+1}) = \mathbf{M}_i\mathbf{x}^t(t_i) + \boldsymbol{\eta}(t_i), \tag{4}$$

where $\boldsymbol{\eta}$ is a noise process with zero mean and covariance matrix $\mathbf{Q}_i$. Then the forecast step for the error covariances is:

$$\mathbf{P}^f(t_{i+1}) = \mathbf{M}_i \mathbf{P}^a(t_i) \mathbf{M}_i^T + \mathbf{Q}_i \,. \tag{5}$$

The equations governing the atmosphere's evolution are nonlinear, so this theoretically optimal Kalman filter cannot be used. The nonlinearity will make pdfs non-Gaussian, not defined by their mean and covariance, but, as explained in the introduction, we cannot afford to integrate the Fokker-Planck equation.

As long as the evolution of errors stays close to linear, pdfs should stay close to Gaussian, and the extended Kalman filter (EKF) should be a good approximation to the evolution of their mean and covariance. Nonlinear operators are used in (1) and (3) for the manipulation of the means of the pdfs, giving

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{P}^f(t_i)\mathbf{H}_i^T \left(\mathbf{H}_i \mathbf{P}^f(t_i) \mathbf{H}_i^T + \mathbf{R}_i\right)^{-1} \left(\mathbf{y}_i^o - H_i\left(\mathbf{x}^f(t_i)\right)\right) \tag{6}$$

$$\mathbf{x}^f(t_{i+1}) = M_i\left(\mathbf{x}^a(t_i)\right) \tag{7}$$

Note that the nonlinear operators H_i and M_i should give the best estimate for the mean of the pdf; they are not generally the same as those which predict any individual state. The operators $\mathbf{H}_i$ and $\mathbf{M}_i$ used in (2) and (5), for the manipulation of the error covariances, are linearized about the trajectory calculated by H_i and M_i. It is clear from (5) that they should be designed to give the best (Gaussian) estimate of the whole pdf of possible errors, not just infinitessimal perturbations.

It is a standard result that the end result of a linear Kalman filter process is identical to that given by a simultaneous best fit to the same observations. We use this to justify 4D-Var methods which work without the explicit manipulation in (2) and (5) of covariance matrices which are too big even to store in current computers. For nonlinear problems 4D-Var is not exactly equivalent to the EKF, since different linearisation states are used, as explained in section 4. But the approximations in deriving nonlinear 4D-Var are very similar to those in the EKF.

3. NWP error covariances - three-dimensional aspects

(a) The effect of covariances on observation weights

Although they have some value in their own right, quantifying likely errors in forecasts for users, the most important use of the error covariances $\mathbf{P}^f(t_i)$ is in (6), determining the weights given to observations. The effect of the variance and correlation are easiest to understand for the simplest case of a single observation of a model variable, in a homogeneous covariance field. The weight at the observation position is $1/\left(1 + R_i / P_i^f\right)$, where R_i is the observation's error variance, and P_i^f is the forecast error variance at the

observation position, i.e. the diagonal element of $\mathbf{P}^f(t_i)$ corresponding to the observed variable. The weight at other positions drops off proportionally to the forecast error correlation (Gandin 1963).

A more important property of (6) is much less intuitive. Observations can complement each other, so that together they have more value than individually. For instance Lorenc (1981, table 1) showed that temperature and wind observations, together with hydrostatic and geostrophic assumptions, determine the vertical and horizontal gradients of the geopotential height field. If suitably place with respect to an accurate geopotential height observation, they greatly enhance its utility by spreading the information in space. This effect is due to the forecast error covariances, which combine in (6) to increase the weights for the synergistic data. In an analogous way, sequential observations of the same field can combine with the knowledge of the evolution equations (embodied in $\mathbf{M}_i$) to give more information about the dynamical fields. For instance sequential observations of a tracer field can give information about the advecting wind (Daley 1996). In the extended Kalman filter this is achieved by using (2) to calculate information about the covariances in $\mathbf{P}^a(t_i)$ between tracer and wind field errors after the first observation, evolving this with (5) to give $\mathbf{P}^f(t_{i+1})$ containing the information, and then using the evolved covariances when using the second tracer observation in (6) at a later time. It would be easy to lose this sequence of interaction in any approximation to (2) and (5). Since the objective of data assimilation is to get as much information as possible from the observations, it is important to design methods which can effectively use the synergistic information in sequential observations; the 4D-Var method described later in this paper does this.

(b) Modelling NWP error covariances in 3D-Var

For NWP models, (2) and (5) are far too large to evaluate. Moreover, the input covariances to the calculations, such as $\mathbf{Q}_i$, are very poorly known. Since matrices $\mathbf{P}$ and $\mathbf{Q}$ can have 10^{14} elements for modern large-scale models, their full structure cannot be determined from empirical validation of actual forecasts; not enough forecasts are made, and accurate validation data does not exist. However we do know about atmospheric dynamics and physics, and can deduce much about their structure from this knowledge. We use this to build a model of typical error covariances, with sufficiently few free parameters that we can determine them empirically, from actual forecasts. It is usual to denote forecast error covariances based on such a model by $\mathbf{B}_{(x)}$, rather than the Kalman filter's $\mathbf{P}$. The subscript indicates which variable's covariance is represented; in this case

$$\mathbf{B}_{(x)} = \left\langle \left(\mathbf{x}^f - \mathbf{x}^t\right)\left(\mathbf{x}^f - \mathbf{x}^t\right)^T \right\rangle, \tag{8}$$

where $\langle\ \rangle$ indicates the expected value, $\mathbf{x}^t$ the true value, and we assume biases are negligible.

Incremental three-dimensional variational assimilation (3D-Var) finds the increment $\delta\mathbf{x}$ to a guess field $\mathbf{x}^g$ which minimises

$$J(\delta\mathbf{x}) = \tfrac{1}{2}\left(\delta\mathbf{x} - \delta\mathbf{x}^b\right)^T \mathbf{B}_{(\mathbf{x})}^{-1}\left(\delta\mathbf{x} - \delta\mathbf{x}^b\right) + \tfrac{1}{2}\left(\mathbf{y} - \mathbf{y}^o\right)^T \mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{y}^o\right), \tag{9}$$

where $\delta\mathbf{x}^b$ is the difference between the background forecast $\mathbf{x}^f$ and the guess $\mathbf{x}^g$. As everything is at the same time, we have dropped for the moment the t_i. In 3D-Var it is usual to define $\mathbf{x}^g$ to be equal to $\mathbf{x}^f$, so $\delta\mathbf{x}^b$ is zero; the distinction becomes important in 4D-Var with an outer-loop, as discussed in section 4(b). The model estimate of the observed values is given by

$$\mathbf{y} = H\left(\mathbf{x}^g\right) + \mathbf{H}\delta\mathbf{x}. \tag{10}$$

It is well known (e.g. Lorenc 1986) that, if $\mathbf{B}_{(\mathbf{x})}$ equals $\mathbf{P}$, and $\mathbf{x}^g$ equals $\mathbf{x}^f$, the solution of (6) is given by the increments which minimise (9):

$$\mathbf{x}^a = \mathbf{x}^g + \delta\mathbf{x}. \tag{11}$$

The big advantage of the variational approach comes when we exploit some physical understanding of the covariances, and transform to a control variable with a simpler covariance matrix.

The traditional arguments for building a model of forecast error covariances relies on the observation that, in the extra-tropics, the true atmosphere and the forecast of it are both close to non-divergence, geostrophic balance and hydrostatic balance (e.g. Lorenc 1981). So the errors should have the same properties. This concept of an approximate slow manifold (Daley 1991 p301) can also be explained in terms of the nonlinear atmospheric system, and chaos theory. Some perturbations grow rapidly, towards an attractor which is in some sense balanced, while most perturbations, such as unbalanced gravity waves, do not grow. So eventually the balanced perturbations dominate the error statistics. If we assume that the balance properties are exact, then we eliminate all gravity waves, and $\mathbf{B}_{(\mathbf{x})}$ has a null-space. (Even if we allow some errors in gravity waves, their amplitudes, and the corresponding eigenvalues in $\mathbf{B}_{(\mathbf{x})}$, are likely to be small.) It is the existence of this null-space in the covariances which allows wind and temperature observations to combine synergistically, as discussed in the previous section. In the statistical analysis method of Lorenc (1981) this was achieved by modelling the elements of $\mathbf{B}_{(\mathbf{x})}$ using non-divergence, geostrophic balance and hydrostatic balance assumptions, and solving (6). In 3D-Var a common approach is to use a transform which block-diagonalises $\mathbf{B}_{(\mathbf{x})}$. Physical arguments are used to choose transformed variables which represent the balanced flow, and gravity waves. Ignoring for the moment moisture, a simple approach is to use perturbations to streamfunction ψ, velocity potential χ, and unbalanced

pressure $\mathbf{b}$ (Parrish and Derber 1992, Lorenc *et al.* 2000). Combining these into a vector $\mathbf{v}$, we know how to define a parameter transformation $\mathbf{U}_p$ such that

$$\delta\mathbf{x} = \mathbf{U}_p \begin{pmatrix} \psi \\ \chi \\ \mathbf{b} \end{pmatrix} = \mathbf{U}_p \mathbf{v}. \tag{12}$$

$$\mathbf{B}_{(\mathbf{x})} = \mathbf{U}_p \mathbf{B}_{(\mathbf{v})} \mathbf{U}_p{}^T. \tag{13}$$

We choose these variables, and construct the transform, such that errors in ψ χ and $\mathbf{b}$ are uncorrelated. Then

$$\mathbf{B}_{(\mathbf{v})} = \begin{pmatrix} \mathbf{B}_{(\psi)} & 0 & 0 \\ 0 & \mathbf{B}_{(\chi)} & 0 \\ 0 & 0 & \mathbf{B}_{(\mathbf{b})} \end{pmatrix}, \tag{14}$$

which is much easier to invert. I delay discussing changes to $\mathbf{x}^g$ until the next section, defining it equal to $\mathbf{x}^f$. Then the transformed variational problem is finding $\mathbf{v}$ which minimises

$$J(\mathbf{v}) = \tfrac{1}{2}\mathbf{v}^T \mathbf{B}_{(\mathbf{v})}^{-1} \mathbf{v} + \tfrac{1}{2}\left(\mathbf{y} - \mathbf{y}^o\right)^T \mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{y}^o\right), \tag{15}$$

$$\mathbf{y} = H\left(\mathbf{x}^f\right) + \mathbf{H}\mathbf{U}_p \mathbf{v}, \tag{16}$$

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{U}_p \mathbf{v}. \tag{17}$$

In practice further horizontal and vertical transformations are used (e.g. Lorenc *et al.* 2000); this simplified presentation is sufficient for the discussions here.

As long as (12) and (14) are valid, (15) gives identical results to (9). If errors forecast errors are close to balance and non-divergence, then $\mathbf{B}_{(\chi)}$ and $\mathbf{B}_{(\mathbf{b})}$ will be small, and the analysis increments will be close to balance and non-divergence. If they are assumed to be zero then we can omit these variables from the transform, giving

$$\delta\mathbf{x} = \mathbf{U}_p \delta\boldsymbol{\psi} = \mathbf{U}_p \mathbf{v} .$$

(18)

$\mathbf{U}_p$ calculates balance model increments $\delta\mathbf{x}$ for all variables from the $\boldsymbol{\psi}$ control variable, pressure and temperature increments coming via the inverse balance equation. Le Dimet and Talagrand (1986) call the omission of $\boldsymbol{\chi}$ and $\mathbf{b}$ "reducing the control variable". They explain that, rather than solving a variational problem with a "weak" constraint given large weight so as to keep it very close to a relationship such as balance, it is easier to solve the problem with a reduced control variable where the relationship is imposed as a strong constraint. With the strong constraint the variational method is implicitly weighting observations in the same way as the simple examples from Lorenc (1981) discussed in 4(a). Balanced wind increments are calculated to be consistent with pressure and temperature observation. In practice, because the transform $\mathbf{U}_p$ improves the conditioning (Lorenc 1997), it is not too difficult to include the extra control variables allowing deviations from balance. Then if for instance $\mathbf{B}_{(\mathbf{b})}$ were to be large, the $\mathbf{b}$ control variable would be varied to fit wind and temperature observations, and there would be no balanced wind increments.

4. Four-dimensional variational data assimilation

(a) Modelling NWP error covariances in 4D-Var

Lets consider a finite time interval covering n cycles of the Kalman filter, and define a set of four-dimensional vectors denoted by underlining, e.g.

$$\underline{\mathbf{x}} = \begin{pmatrix} \mathbf{x}(t_0) \\ \mathbf{x}(t_1) \\ \vdots \\ \mathbf{x}(t_{n-1}) \end{pmatrix}.$$

(19)

Then, with appropriate extensions to the interpolations in H and $\mathbf{H}$, the four-dimensional versions of (9) and (10) have identical form:

$$J(\delta\underline{\mathbf{x}}) = \tfrac{1}{2}\left(\delta\underline{\mathbf{x}} - \delta\underline{\mathbf{x}}^b\right)^T \mathbf{B}_{(\underline{\mathbf{x}})}^{-1}\left(\delta\underline{\mathbf{x}} - \delta\underline{\mathbf{x}}^b\right) + \tfrac{1}{2}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)^T \mathbf{R}^{-1}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right),$$

(20)

$$\underline{\mathbf{y}} = H\left(\underline{\mathbf{x}}^g\right) + \mathbf{H}\delta\underline{\mathbf{x}} .$$

(21)

The $\delta\underline{\mathbf{x}}^{a}$ which minimises (20) give a simultaneous fit to all the observations over the n cycles. This approach, or the corresponding non-incremental minimisation of $J(\underline{\mathbf{x}})$, is optimal as long as $\mathbf{B}_{(\underline{\mathbf{x}})}$ properly describes the errors in $\underline{\mathbf{x}}^{b}$. This requires the Gaussian approximation of the pdfs to be valid, which needs the forecast equations to be only weakly nonlinear. Since NWP models are nonlinear the Gaussian approximation is not necessarily optimal. An advantage of the incremental approach is that we can conceptually filter the perturbations to retain scales whose evolution is close to linear; for these (20) is close to optimal. Of course direct minimisation of (20) is not practicable for NWP models; $\mathbf{B}_{(\underline{\mathbf{x}})}$ is too large, and not sparse. So we use the same approach as in (12) and (14), and use a transformed variable with a simpler block-diagonal covariance structure. To emphasise the analogy with the 3D-Var transform that used the geostrophic balance equation, I use $\underline{\mathbf{v}}$ for the variable and $\underline{\mathbf{U}}$ for the transform, but in this case $\underline{\mathbf{U}}$ is based on the recursive application of the tangent-linear equivalent of prognostic equation (4):

$$\delta\mathbf{x}(t_{i+1}) = \mathbf{M}_i \delta\mathbf{x}(t_i) + \boldsymbol{\eta}(t_i).$$

(22)

The initial $\delta\mathbf{x}(t_0)$, together with estimates of the stochastic errors at each step of the forecast, constitute the transformed control variable $\underline{\mathbf{v}}$, from which we can calculate the full $\delta\underline{\mathbf{x}}$:

$$\delta\underline{\mathbf{x}} = \underline{\mathbf{U}} \begin{pmatrix} \delta\mathbf{x}(t_0) \\ \boldsymbol{\eta}(t_0) \\ \vdots \\ \boldsymbol{\eta}(t_{n-2}) \end{pmatrix} = \underline{\mathbf{U}}\underline{\mathbf{v}}.$$

(23)

Then it is possible to write a matrix form for $\underline{\mathbf{U}}$ (in practical computations (22) is used):

$$\underline{\mathbf{U}} = \begin{pmatrix} \mathbf{I} & 0 & \cdots & 0 \\ \mathbf{M}_0 & \mathbf{I} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}_{n-2}\cdots\mathbf{M}_0 & \mathbf{M}_{n-2}\cdots\mathbf{M}_1 & \cdots & \mathbf{I} \end{pmatrix}.$$

(24)

Then, in line with the stochastic assumption of the Kalman filter, the covariance of the transformed control variable is block diagonal:

$$\mathbf{B}_{(\underline{\mathbf{v}})} = \begin{pmatrix} \mathbf{P}^f(t_0) & 0 & \cdots & 0 \\ 0 & \mathbf{Q}_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Q}_{n-2} \end{pmatrix}.$$

(25)

We can use this in the 4D-Var equivalent of (15) (16) and (17). (In practice we would use (13), and further transformations from 3D-Var, to diagonalise $\mathbf{P}^f(t_0)$.) Equations (24) and (25) together define the four-dimensional covariance, which fortunately we do not need to evaluate:

$$\mathbf{B}_{(\underline{\mathbf{x}})} = \underline{\mathbf{U}}\mathbf{B}_{(\underline{\mathbf{v}})}\underline{\mathbf{U}}^T.$$

(26)

With a control variable defined in this way, and a linear model, 4D-Var is equivalent to the fixed-interval Kalman smoother (Ménard and Daley 1996). The Kalman filter adds information from each batch of observations sequentially, so is it only equivalent to 4D-Var at the end, when all the information has been added. Incremental 4D-Var is not exactly equivalent to the extended Kalman filter, because of differences between the nonlinear trajectories about which the tangent-linear operators are defined. I postpone further discussion of nonlinearities, and the trajectory used in incremental 4D-Var, until the next sub-section on the outer-loop.

Experimental 4D-Var systems using model error as a control variable as in (23) have been developed by Derber (1989) and Zupanski (1997). One problem with (25) is that real model errors are correlated in time; for instance errors in a parametrization scheme of cloud processes are likely to be similar as long as the cloud persists. A more practical problem is that having independent error variables at each model time-step makes control variable $\underline{\mathbf{v}}$ very large. Both can be ameliorated by assuming a model for the time correlation of errors, and making the independent stochastic variables in this model the new control variables. The simplest approach is to assume model errors are constant over a period, so we have a single error control variable $\boldsymbol{\eta}$ which is use in (22) for all the time-steps in the period. (In fact a similar assumption is made in the Kalman filter when observations are batched for periods longer than a model time-step, and a single $\mathbf{Q}_i$ is used for each batch).

It is possible to derive equivalent algorithms which solve a variation problem with control variables corresponding to each observation, rather than the model variables I use here (Bennett 1992, Courtier 1997). These can allow for model errors without explicitly estimating them, and will be computationally more efficient when the number of observations is small. I will not discuss these further because the focus of this paper is on the error covariances.

It is impossible to get an accurate estimate of the error statistics for a single time-step of a model - there is insufficient validation data. However some idea of their characteristics can be obtained in diagnostic studies (e.g. Klinker and Sardeshmukh 1992). These show persistent systematic errors, due to deficiencies in parametrizations of sub-grid-scale

heating and drag processes, which at source are not in geostrophic balance. But the cumulative errors from these terms will evolve and grow as discussed in section 3(a), so the net result after some hours is errors which are largely balanced. This effect may be enhanced by the inclusion (deliberate or not) in the NWP model of diffusion terms which damp unbalanced modes more rapidly than in the real atmosphere. Hence a control variable $\boldsymbol{\eta}_i$ meant to represent the cumulative model errors over a period of several hour must have a covariance matrix $\mathbf{Q}_i$ with some of the balanced properties discussed in section 3(b). Beyond these qualitative arguments, we have few means of quantifying the size or structure of $\mathbf{Q}_i$, or of any further time correlation with errors in an adjacent period.

Mainly for reasons of expense and complexity, operational 4D-Var schemes to date have not allowed for model error, using only the initial perturbation $\delta\mathbf{x}(t_0)$ as control variable. In a manner analogous to the early statistical interpolation schemes (e.g. Lorenc 1981) that omitted divergence from the covariance model, they are omitting the variable which allows deviations from the model equations, and hence are imposing the perturbation model equations $\mathbf{M}_i$ as a strong constraint on the four-dimensional increments $\delta\underline{\mathbf{x}}$. Just as later 3D-Var methods went on to include a divergence control variable, I expect more sophisticated 4D-Var systems in future to include an allowance for model errors, in order to be able to extend the assimilation window beyond the limit where the model errors can be neglected. However, given the lack of knowledge of the size or structure of $\mathbf{Q}_i$, only rather simple methods are justified.

(b) Incremental 4D-Var with an outer-loop

In incremental 4D-Var for a nonlinear model, it is usual to consider an outer-loop, where the incremental minimisation of (20) is repeated for a succession of trajectories (Courtier *et al.* 1994). The unified notation of Ide *et al.* (1997) uses the superscript ^{g} to denote current best-guess values which are held constant during each variational minimisation. It is also convenient to define $\underline{\mathbf{v}}^g$ as the sum of increments from previous loops; initially zero. An outer-loop algorithm including model error control variables is based on a current best estimate of the four-dimensional trajectory $\underline{\mathbf{x}}^g$ evaluated using the nonlinear model and the elements of $\underline{\mathbf{v}}^g$:

$$\begin{aligned}\mathbf{x}^g(t_0) &= \mathbf{x}^f(t_0) + \delta\mathbf{x}^g(t_0)\\ \mathbf{x}^g(t_{i+1}) &= M_i\left(\mathbf{x}^g(t_i)\right) + \boldsymbol{\eta}^g(t_i) \quad \textit{for } i = 0, n-1\end{aligned} \tag{27}$$

(We cannot use a matrix representation since M_i is nonlinear.) The tangent-linear operators $\mathbf{H}_i$ and $\mathbf{M}_i$ are linearized about $\underline{\mathbf{x}}^g$. The variational problem is to minimise

$$J(\underline{\mathbf{v}}) = \tfrac{1}{2}\left(\underline{\mathbf{v}}^g + \underline{\mathbf{v}}\right)^T \mathbf{B}^{-1}_{(\underline{\mathbf{v}})}\left(\underline{\mathbf{v}}^g + \underline{\mathbf{v}}\right) + \tfrac{1}{2}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)^T \mathbf{R}^{-1}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right), \tag{28}$$

where the increments $\delta\underline{\mathbf{x}}$ are calculated each iteration from the control variable using (23) and (24), and the model estimates of the observations ($\underline{\mathbf{y}}$) are calculated using (21). When the $\underline{\mathbf{x}}^{g}$ which minimises (28) has been found, the guess for the next outer iteration is given by

$$\underline{\mathbf{v}}^{g+} = \underline{\mathbf{v}}^{g} + \underline{\mathbf{v}}. \tag{29}$$

We call the iterative minimisation of (28) the inner-loop, and the updating of the guess using (29) and (27) the outer-loop. Note that if this outer-loop iteration converges[2], then increments $\delta\underline{\mathbf{x}}$ in the final variational minimisation will be zero, and the model prediction of the observations in (21) will depend only on the form of the full nonlinear model M_i and observation operators H_i, rather than the linear operators $\mathbf{M}_i$ and $\mathbf{H}_i$. Just as in the extended Kalman filter, the linearized operators are only used in the covariance manipulations which determine the weight given to the innovations $\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^{o}\right)$, rather than in their evaluation. Since, as discussed in 3(b) and 4(b) above, we have only very limited data about the covariances being manipulated, it is reasonable to make simplifications in the linear operators. This can make 4D-Var affordable, and was the reason for introducing the incremental approach (Courtier *et al.* 1994).

(c) Incremental 4D-Var with a simplification operator

For an NWP model the most obvious simplification is a reduction in resolution, but it is also useful to include simplifications to the representation of physical processes such as cloud. We introduce a simplified mode state $\mathbf{w}$, and a simplification operator S that is usually linear, but may have nonlinear terms, for instance in the treatment of moisture variables. We are assuming that the desired analysis can be expressed in term of increments to this state:

$$\begin{aligned} \mathbf{w}^{g} &= S\left(\mathbf{x}^{g}\right) \\ \mathbf{w} &= S(\mathbf{x}) \\ \delta\mathbf{w} &= S(\mathbf{x}) - S\left(\mathbf{x}^{g}\right) \end{aligned} \quad . \tag{30}$$

To use the simplified increment we need to be able to add it to $\mathbf{x}^{g}$:

$$\mathbf{x} = \mathbf{x}^{g} + S^{-1}\delta\mathbf{w}. \tag{31}$$

We can now proceed as above, replacing (23) by

[2] Criteria for convergence are discussed in the appendix of Veersé and Thépaut (1998)..

$$\delta \underline{\mathbf{w}} = \underline{\mathbf{U}} \begin{pmatrix} \delta \mathbf{w}(t_0) \\ \eta(t_0) \\ \vdots \\ \eta(t_{n-2}) \end{pmatrix} = \underline{\mathbf{U}} \underline{\mathbf{v}} .$$

(32)

As S is a simplification operator it does not have an inverse; we use the generalised inverse S^{-1}, which we will call the incrementing operator. When S is just a decrease in resolution, S^{-1} can set increments for the unresolved scales to zero, or alternatively can interpolate from a low-resolution to a high-resolution grid. Modern NWP models often have a sophisticated representation of microphysical processes, with independent variables to represent water vapour, cloud liquid water, cloud ice, etc.. These add to the cost of the model, and, given other uncertainties; it is unlikely that their inclusion will significantly improve predictions of the evolution of error covariances. So we should allow for simplification of this representation, for instance including in $\delta \mathbf{w}$ just increments to a single total-moisture variable.

We can apply S to each time in the four-dimensional trajectory $\underline{\mathbf{x}}$, giving a simplified trajectory $\underline{\mathbf{w}}$. We define a simplified perturbation forecast model $\widetilde{\mathbf{M}}_i$, linearized about $\underline{\mathbf{w}}$, such that

$$\delta \mathbf{w}(t_{i+1}) = \widetilde{\mathbf{M}}_i \delta \mathbf{w}(t_i) + \eta(t_i) .$$

(33)

This defines increments to the full trajectory, using

$$\mathbf{x}^g(t_0) = \mathbf{x}^f(t_0) + S^{-1} \delta \mathbf{w}(t_0)$$
$$\mathbf{x}^g(t_{i+1}) = M_i\left(\mathbf{x}^g(t_i)\right) + S^{-1} \eta(t_i) \quad for \;\; i = 0, n-1$$

(34)

Although the simplification has in practice been necessary for computational reasons, it also has a role in filtering motions whose evolution is very nonlinear over the time-window.

Lorenc *et al.* (2000) split the observation operator H into an interpolation L of the model variables to the time and position of the observation, and an observation operator V which converts the model variables into those observed. For instance in the case of a satellite sounding the output of L will be a column of model variables at the position of the observations, and V will be the radiative transfer calculations (Eyre *et al.* 1993). Then the predictions of the observations from the incremented model are given by

$$\underline{\mathbf{y}} = V\left(L\left(\underline{\mathbf{x}}^g\right) + S_v^{-1} \widetilde{L} \delta \underline{\mathbf{w}}\right).$$

(35)

Here S_v^{-I} denotes only that part of the incrementing operator which deals with the simplified variables; the change of resolution is allowed for by the interpolation $\tilde{L}$ working directly from the simplified grid. For instance if the full model carries independent variables for water vapour, cloud liquid water, cloud ice, etc., then S_v^{-I} has to specify how to increment these using the simplified representation such as a total water perturbation. To achieve this, S_v^{-I} can be designed using a physically based parametrization of the relationships between the full variables. (35) can then be used in a variational inner-loop like (28).

An advantage of the variational method is that it allows direct use of observations not related to the model variables (e.g. Eyre *et al.* 1993). We can retain this advantage by allowing V S_v^{-I} and $\tilde{L}$ to be nonlinear functions of $\delta\underline{\mathbf{w}}$. This makes (28) non-quadratic; since it is derived by taking the logarithm of the posterior pdf (Lorenc 1986), the pdf is non-Gaussian. Alternatively one can get a completely quadratic inner-loop by using

$$\underline{\mathbf{y}} = V\left(L\left(\underline{\mathbf{x}}^g\right)\right) + \mathbf{V}\mathbf{S}_v^{-I}\tilde{\mathbf{L}}\delta\underline{\mathbf{w}}. \qquad \textbf{(36)}$$

This allows more efficient minimisation methods to be used. The outer-loop can still allow for nonlinearities. It is the form (36) we use when considering the implicit covariances.

The simplification can be regarded as another reduction of the control variable. Where it is reducing the resolution, then we are assuming the small scales which are omitted in the simplified state have negligible variance. Where it simplifies the variables, we must specify in S_v^{-I} relationships giving increments to the full set of variables from the simplified variables. Eliminating the simplified variables then gives implied incremental relationships between some of the full variables; we are implicitly using these as a strong constraint. Our model for the full three-dimensional covariances is thus

$$\mathbf{B}_{(\mathbf{x})} = \mathbf{S}_v^{-I}\mathbf{B}_{(\mathbf{w})}\mathbf{S}_v^{-T}. \qquad \textbf{(37)}$$

Since the simplification operator is rank-deficient, the covariance matrix defined by (37) has a null-space corresponding to the small scales, and deviations from the implicit inter-variable relationships described above.

5. Discussion and conclusions

I have started from the Bayesian assumption that the optimal model state (called the analysis) is defined by the posterior pdf, given by combining pdfs describing the observed information, and the prior pdf describing information in a forecast model state. Given stochastic equations for the time-evolution of an atmospheric state using a model with errors, it is in principle possible to calculate the time-evolved pdf. However this is impracticable for a large problem like NWP; the object of this paper is to derive the incremental 4D-Var method as an approximation to the optimal method, and in so doing

give some insight into the important characteristics which make 4D-Var a good approximation.

Most of the approximations are derived by modelling the pdfs as multivariate Gaussian distributions. A pdf can then be defined by its mean, and a covariance describing the perturbations about the mean. The major point of my presentation of incremental 4D-Var as being equivalent to the Kalman filter is that we calculate the time-evolution of the mean state with the full nonlinear NWP model, while the simplified perturbation forecast model is only used to calculate the evolution of the covariances.

4D-Var is often presented as a least squares best fit to the observations, constrained to be consistent with the forecast model (e.g. Talagrand and Courtier 1987), or as the search for the most probable analysis: the mode of the pdf (e.g. Lorenc 1986). Descent algorithms use the local derivatives of the pdf to find the maximum. For Gaussian pdfs the approaches are identical. But the atmosphere evolves nonlinearly, and exhibits chaotic behaviour at all scales. Some perturbations grow rapidly; the growth being most rapid for small scales. Prediction to longer times is only possible because the small-scale growth saturates[3] at a sufficiently low amplitude that the larger scales are not obscured. This means that a realistic NWP model may represent all the scales, relying on the nonlinear saturation effects to limit chaotic growth of small scales. But the exact tangent linearisation of such a model will not exhibit the same saturation; perturbations will continue to grow at the rate of infinitessimal perturbations, which for small scales will be large enough rapidly to obscure the larger-scale perturbations in which we are more interested. This effect will occur even if the nonlinear model does not contain non-differentiable threshold algorithms. If the NWP model represents scales which behave chaotically over the time-window being considered, then very small changes to the initial conditions may make large changes to the probabilities, and finding the maximum of the pdf by a descent algorithm is impossible. To avoid this problem, practical 4D-Var algorithms use reduced resolution and simplified "regularised" parametrization algorithms in their linear models, which are thus not tangent to the best full model. However often these are regarded as approximations to the ideal tangent-linear model. By seeking the mean rather than the mode of the pdf, the approach proposed in section 4 avoids these problems. The linear model should provide a good approximation to the evolution of finite perturbations; in general it will not be the tangent-linear model obtained by automatic differentiation of the full model. To emphasise this I prefer to call it the perturbation forecast model.

A problem common in Bayesian methods is a serious lack of knowledge about the prior pdf. In NWP assimilation this is compounded by the chaotic nature of the atmosphere, which means that even if the variance of model error source terms $\mathbf{Q}_i$ is small, some errors may grow rapidly. So we cannot rely on the accumulation of observed information over many data assimilation cycles to reduce the error in our model state, the variances of the growing model errors will become significant. Since we know little about model errors, as discussed in 4(b), we can only ever have a limited knowledge of the forecast error covariance $\mathbf{P}^f$. It is therefore not justified to expend a large effort calculating its evolution; simplifications introduced in a perturbation forecast model to save

[3] It must saturate since climatological variability is observed to be bounded.

computational effort may not introduce significant errors in our knowledge of the covariances.

In section 4 I derived 4D-Var from a four-dimensional variational problem, with a four-dimensional control variable with prior errors described by covariance $\mathbf{B}_{(\underline{\mathbf{x}})}$. The linear model $\mathbf{M}_i$ enters only as a way of modelling the covariances $\mathbf{B}_{(\underline{\mathbf{x}})}$. Other limitations (discussed above) to our knowledge of forecast error covariances then justify the use of a simplified perturbation forecast model $\tilde{\mathbf{M}}_i$. This derivation of 4D-Var allows for model errors, just as the Kalman filter does. Seen this way, it is incorrect to say that the "perfect model assumption" is a fundamental characteristic of 4D-Var. It is show to be a simplifying assumption in the error covariance model, analogous to the non-divergence assumption used in early three-dimensional statistic interpolation methods. Although the perfect model assumption is used in currently operational schemes, it is likely that as 4D-Var methods become more mature it will be relaxed, just as current 3D-Var methods no longer assume non-divergence in their covariance modelling.

The incremental 4D-Var algorithm makes approximations in assuming Gaussian pdfs (slightly relaxed in the nonlinear form (35)), in the simplification operator, and in the forecast error covariance at initial time ($\mathbf{P}^f(t_0)$, which is modelled as in 3D-Var), but otherwise it remains close to the optimal Bayesian analysis. In particular, as long as the simplified model correctly represents the large-scale dynamical equations in whose accuracy we have confidence, then the null-space of $\mathbf{B}_{(\underline{\mathbf{x}})}$ is preserved. I suggest that this property of the term involving $\mathbf{B}_{(\underline{\mathbf{x}})}^{-1}$ in (20), penalising solutions which are inconsistent with the dynamical equations when fitting the observation in the second term, is one of the main benefits of the 4D-Var method. It allows dynamically consistent information to be extracted from a time-sequence of observations, in an analogous manner to the extraction of dynamically consistent linking of temperatures and winds in methods using three-dimensional multivariate covariances modelled using balance equations.

The interpretation of the properties of 4D-Var through the implicit error covariances is not new (e.g. Thépaut *et al.* 1993, Thépaut *et al.* 1996). These authors make the point that the error covariances at time t_j are given by $\mathbf{M}_{j-1}...\mathbf{M}_0\mathbf{P}^f(t_0)\mathbf{M}_0^T...\mathbf{M}_{j-1}^T$ (making the perfect model assumption and ignoring terms in $\mathbf{Q}_i$). So, even though it is usual to approximate $\mathbf{P}^f(t_0)$ by an averaged covariance model which does not take account of the flow pattern at time t_0, the covariances at time t_j do depend on the flow patterns, through the linearized model $\mathbf{M}$. This is visible in the three-dimensional covariance patterns, which (as long as the evolution time t_j is long enough) tend to follow baroclinic zones in a plausible way (Thépaut *et al.* 1996 fig 12). Moreover the covariances between an observation at time t_j and the control variables at time t_0 reflect the growth of perturbations between these times (Thépaut *et al.* 1993, figures 4 & 2). This is usually illustrated, as in these examples, using observations at a single time. This gives a correct and useful insight into the implicit covariances in 4D-Var. However, just as implicit balance constraints in 3D-Var are not immediately apparent from the mass-wind covariance patterns, the 4D-Var covariances do not immediately make clear the effect of the model constraints in the time dimension. This effect does not need a 4D-Var

assimilation window long enough for covariances to evolve into the synoptically plausible structures discussed by Thépaut *et al.* (1993, 1996), as long as the window is long enough for observed tendencies to be large compared to observational errors.[4]

The incrementing operator $\mathbf{S}^{-1}$ can play a very similar role in the covariance model to the transform $\mathbf{U}$, in allowing a reduced control variable to determine a higher resolution, less simple model state. Implicit in the design of $\mathbf{S}^{-1}$ are relationships between the full model variables and the simplified variables which imply relationships between increments to different full model variables. The main advantage of simplifying, rather than reducing the control variable earlier in the variable transform, is in the reduced design and computational costs of the perturbation forecast model $\tilde{\mathbf{M}}_i$.

The main NWP model in the extended Kalman filter (7) and 4D-Var (27) is used to calculate the evolution of the mean of the pdf. An advantage of the averaging is that threshold processes, like the onset of precipitation or convection, will occur at different times in each realisation. Hence the evolution of the mean state should be smooth and differentiable, making the tangent-linear model well defined. Since the model is nonlinear, it is not in general true that the same model can forecast the evolution of any single realisation. This is particularly true when the NWP model is capable of resolving structures whose detail is uncertain, such as turbulence and convection. The mean state will average out these structures, yet its evolution should be affected by the average nonlinear fluxes from the hypothetical ensemble of states with different realisations of the turbulent structures. Constructing parametrizations to do this is conceptually difficult, not least because the filtering properties of the averaging process will depend on the spread of the pdf; so the parametrization should depend on this and not just the mean state. The problems are not just seen in 4D-Var, they are fundamental to deterministic NWP. If the error variance is significant, the single best estimate will not be very useful without considering it. For instance in a convective airmass, if there are insufficient observations to resolve where the convective cells are, the parametrizations must allow for their effect, even if the model has enough resolution to represent them, since their structure will be absent from the "best estimate" initial conditions. But in a better observed region, for instance near a radar, the same model can explicitly represent the convection, and parametrization is not needed. There would be several practical problems in designing the model and data assimilation system for an NWP system like this, as well as in interpreting its output. An alternative is to use an ensemble to represent the range of possible states, each being a plausible realisation with explicitly modelled convection. The model can then be designed to forecast these without needing to know if they match reality (in a well observed region), or are just one possible realisation (in a poorly observed region). A sufficiently large ensemble might sample the pdf well enough to be used in data assimilation. Lorenc (2002) compares the ensemble Kalman filter to 4D-Var using the ideas of this paper.

[4] 4D-Var only makes optimal use of observations within the same variational problem. In a practical data assimilation algorithm consisting of cycles of 4D-Var for batches of observations, precise tendency information between observations in different batches is not retained.

References

Bennett, A. F., 1992: "Inverse Methods in Physical Oceanography. Cambridge University Press, New York. 346pp.

Courtier, P. 1997: "Dual formulation of four-dimensional variational assimilation." *Quart. J. Roy. Met. Soc.*, **123**, 2449-2461

Courtier, P., J-N. Thépaut, and A. Hollingsworth, 1994: "A strategy for operational implementation of 4D-Var, using an incremental approach". *Quart. J. Roy. Met. Soc.*, **120**, 1367-1387.

Daley R. 1996: "Recovery of the one and two dimensional windfields from chemical constituent observations using the constituent transport equation and an extended Kalman filter." *Met. and Atmos. Phys.*, **60**, 119-136.

Daley R. 1991: "Atmospheric Data Analysis". *Cambridge University Press*, ISBN 0-521-38215-7

Dee, Dick P. 1991: "Simplification of the Kalman filter for meteorological data assimilation". *Quart. J. Roy. Met. Soc.*, **117**, 365-384.

Derber, John C. 1989: "A variational continuous assimilation technique". *Mon. Wea. Rev.*, **117**, 2437-2446

Ehrendorfer, M., 1994a: "The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory." *Mon. Wea. Rev.*, **122**, 703-713.

Ehrendorfer, M., 1994b: "The Liouville equation and its potential usefulness for the prediction of forecast skill. Part II: Applications." *Mon. Wea. Rev.*, **122**, 714-728.

Evensen, G., 1994: "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics." *J. Geophys., Res.*, **99**(C5), 10143-10162.

Eyre,J.R., G.A.Kelly, A.P.McNally, E.Andersson, and A Persson 1993: "Assimilation of TOVS radiance information through one-dimensional variational analysis" *Quart.J.Roy.Met.Soc.*, **119**, 1427-1464

Gandin, L.S. 1963: "Objective analysis of meteorological fields." Izd. Leningrad (in Russian), English translation by by the Israel program for scientific translations, 1965, 242pp.

Houtekamer, P. L. and H. L. Mitchell, 1998: "Data assimilation using an ensemble Kalmam filter technique." *Mon. Wea. Rev.*, **126**, 796-811

Ide, K., Courtier, P., Ghil, M., and Lorenc, A.C. 1997: "Unified notation for data assimilation: Operational, Sequential and Variational" *J. Met. Soc. Japan*, Special issue "Data Assimilation in Meteorology and Oceanography: Theory and Practice." **75**, No. 1B, 181--189

Jazwinski, A. H., 1970: "Stochastic processes and filtering theory." *New York Acad. Press*, 376pp.

Klinker, E. and P. D. Sardeshmukh, 1992: "The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements." *J. Atmos. Sci.*, **49**, 608-627

Le Dimet, F-X. and Talagrand, O. 1986: "Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects." *Tellus*, **38A**, 97-110.

Lewis, John M. and Derber, John C. 1985: "The use of adjoint equations to solve a variational adjustment problem with advective constraints". *Tellus*, **37A**, 309-322

Lorenc, A.C. 1981: "A global three-dimensional multivariate statistical analysis scheme." *Mon. Wea. Rev.*, **109**, 701-721.

Lorenc, A.C. 1986: "Analysis methods for numerical weather prediction." *Quart. J. Roy. Met. Soc.*, **112**, 1177-1194.

Lorenc, A.C., 1997: "Development of an Operational Variational Assimilation Scheme." *J. Met. Soc. Japan*, Special issue "Data Assimilation in Meteorology and Oceanography: Theory and Practice." **75**, No. 1B, 339-346.

Lorenc, A. C. 2002: "Modelling of error covariances by the ensemble Kalman filter." *In preparation*

Lorenc, A. C., S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne and F. W. Saunders. 2000: The Met. Office Global 3-Dimensional Variational Data Assimilation Scheme. *Quart. J. Roy. Met. Soc.*, **126**, 2991-3012.

Ménard, R. and R. Daley, 1996: "The application of Kalman smoother theory to the estimation of 4DVAR error statistics." *Tellus*, **48A**, 221-237

Parrish,David F., and Derber, John C. 1992: "The National Meteorological Center's Spectral Statistical-Interpolation analysis system" *Mon.Wea.Rev.* **120,** 1747-1763

Talagrand, O. and Courtier, P. 1987: "Variational assimilation of meteorological observations with the adjoint vorticity equation. Part 1: Theory". *Quart. J. Roy. Met. Soc.*, **113**, 1311-1328

Thépaut, Jean-Noel, Ross N Hoffman, and Philippe Courtier 1993: "Interactions of Dynamics And Observations in a Four-Dimensional Variational Assimilation". *Mon. Wea. Rev.*, **12**, 3393-3414

Thépaut, Jean-Nöel, P. Courtier, G. Belaud and G Lemaître: 1996 "Dynamical structure functions in a four-dimensional variational assimilation: A case study" *Quart. J. Roy. Met. Soc.*, **122**, 535-561

Veersé, F. and J.-N. Thépaut 1998: "Multiple-truncation incremental approach for four-dimensional variational data assimilation". *Quart. J. Roy. Met. Soc.*, **124**, 1889-1908.

Zupanski, D., 1997: "A general weak constraint applicable to operational 4DVAR data assimilation systems." *Mon. Wea. Rev.*, **125**, 2274-2292